

AN INVESTIGATION OF THE CONTENT VALIDITY AND BACKWASH
EFFECT OF THE END-OF-TERM ORAL ASSESSMENT TEST ADMINISTERED
AT HACETTEPE UNIVERSITY, DEPARTMENT OF BASIC ENGLISH (DBE)

A THESIS PRESENTED BY
HASAN ÖSKEN

TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ARTS
IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

BILKENT UNIVERSITY

JULY 1999

PE
1068
.T8
085
1999

AN INVESTIGATION OF THE CONTENT VALIDITY AND BACKWASH
EFFECT OF THE END-OF-TERM ORAL ASSESSMENT TEST ADMINISTERED
AT HACETTEPE UNIVERSITY, DEPARTMENT OF BASIC ENGLISH (DBE)

A THESIS PRESENTED BY

HASAN ÖSKEN

TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF ARTS

IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

BILKENT UNIVERSITY

JULY 1999

Hasan ÖSKEN

tarafından hazırlanmıştır

PE
1068
.T8
O85
1999
B048831

ABSTRACT

Title: An Investigation of the Content Validity and Backwash Effect of the End-of-Term Oral Assessment Test Administered at Hacettepe University, Department of Basic English.

Author: Hasan Ösken

Thesis Chairperson: Dr. William Snyder
Bilkent University, MA TEFL Program

Committee Members: Dr. Necmi Akşit
Michele Rajotte
Bilkent University, MA TEFL Program

This study investigated the content validity and backwash effect of the end-of-term Oral Assessment Test (OAT) administered at Hacettepe University, Department of Basic English. The end-of-term OAT is a final achievement test used to measure students' oral language abilities. The content validity of the OAT was investigated in terms of consistency between the learning goals set for the students in the course book content and taught in the language program and the content of the OAT. A related issue to the content validity was the backwash effect of the OAT, which is the effect of the test on teaching and learning in the classroom.

The idea behind this study originated from overhearing complaints from the teachers and students that the OAT did not test what students had learned in the framework of the course book content. For this reason, I launched this study to investigate the content validity and backwash effect of the OAT.

This study included three groups of subjects: 14 B-level subject teachers and two testers, 62 B-level students and three administrators.

To gather data, questionnaires were given to the three groups of subjects mainly to obtain their opinions about the course book content and the content of the OAT.

Apart from that, the types of speaking tasks in both the course book and the OAT were identified and compared with each other with the aim of revealing consistency.

Data from questionnaires were analyzed using frequencies and percentages and the results were shown in tables. For the comparison of the speaking task types between the course book content and the OAT, the types of speaking tasks specified in the course book content are documented and then matched with those tested in the OAT.

The results of the documentary analysis of the types of speaking tasks both in the course book content and content of the OAT showed that although there were 13 types of speaking tasks occurring in the course book, only three of them were on the OAT. This resulted in a low degree of the content validity of the OAT. The results of the questionnaires supported the findings of the documentary analysis above indicating that the majority of the speaking task types in the course book were not included and tested in the OAT, which proved inconsistency to a certain extent. In addition, through the questionnaires, it was revealed that students did not put a lot of time and effort in the classroom on the types of the speaking tasks which were not tested and were of no value in terms of passing or failing the OAT.

The findings suggest that the content of the OAT should be redesigned to include a greater variety of speaking tasks, such as discussions, role plays, and simulations. Another suggestion is that an oral assessment test should be administered at least twice a term in addition to the one administered at the end of the term.

BILKENT UNIVERSITY
INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES
MA THESIS EXAMINATION RESULT FORM

July, 31, 1999

The examining committee appointed by the Institute of Economics and Social
Sciences

for the thesis examination of the MA TEFL student

Hasan Ösken

has read the thesis of the student.

The committee has decided that the thesis of the student is satisfactory.

Thesis Title: An Investigation of the Content Validity and Backwash Effect of the
End-of-Term Oral Assessment Test Administered at Hacettepe
University, Department of Basic English (DBE)

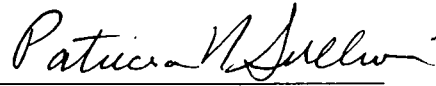
Thesis Advisor: Dr. Patricia Sullivan
Bilkent University, MA TEFL Program

Committee Members: Dr. William E. Snyder
Bilkent University, MA TEFL Program

Dr. Necmi Akpıt
Bilkent University, MA TEFL Program

Michele Rajotte
Bilkent University, MA TEFL Program

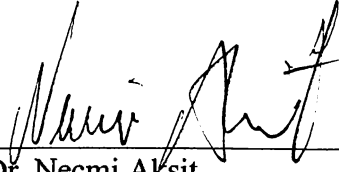
We certify that we have read this thesis and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts.



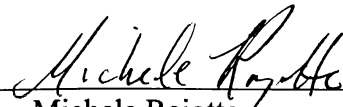
Dr. Patricia Sullivan
(Advisor)



Dr. William E. Snyder
(Committee Member)



Dr. Necmi Akşit
(Committee Member)



Michele Rajotte
(Committee Member)

Approved for the
Institute of Economics and Social Sciences



Ali Karaosmanoğlu
Director
Institute of Economics and Social Sciences

TABLE OF CONTENTS

LIST OF TABLES.....	x
CHAPTER 1	INTRODUCTION..... 1
	Introduction..... 1
	Background of the Study..... 4
	Statement of the Problem..... 5
	Purpose of the Study..... 6
	Significance of the Study..... 6
	Research Questions..... 7
CHAPTER 2	LITERATURE REVIEW..... 8
	Introduction..... 8
	Definition, Purpose and Function of Language Tests..... 8
	Types of Tests..... 11
	Norm-referenced Tests..... 11
	Criterion-referenced Tests..... 12
	Comparison Between Proficiency and Achievement Tests..... 12
	Validity of Language Tests..... 14
	Difficulties of Testing Speaking..... 16
	The Instruments Used for Testing Speaking..... 19
	Role play..... 19
	Discussion..... 20
	Oral Interview..... 20
	Simulation..... 21
	Description..... 21
	Expressing Ideas and Opinions on a Topic..... 22
	Related Parties Involved in an Oral Test..... 24
	Assessment Criteria of an Oral Test..... 25
	Backwash Effect..... 27
CHAPTER 3	METHODOLOGY..... 30
	Introduction..... 30
	Subjects..... 31
	B-level Subject Teachers and Testers..... 31
	B-level Students..... 32
	Administrators of the DBE..... 32
	Materials..... 32
	Procedures..... 34
	Data Analysis..... 35

CHAPTER 4	DATA ANALYSIS.....	37
	Overview of the Study.....	37
	Data Analysis Procedure.....	38
	Results of the Study.....	40
	Analysis of the Speaking Tasks.....	40
	Analysis of the Questionnaires.....	42
	Subject Teachers and Testers Questionnaire.....	42
	B-level Students Questionnaire.....	47
	Administrators Questionnaire.....	51
CHAPTER 5	CONCLUSIONS.....	53
	Summary of the Study.....	53
	Discussion of Findings and Conclusions.....	54
	Limitations of the Study.....	57
	Implications.....	57
	General Implications.....	57
	Institutional Implications.....	58
	Further Research.....	58
REFERENCES.....		59
APPENDICES.....		61
	Appendix A:	
	Teachers and Testers Questionnaire.....	61
	Appendix B:	
	B-level Students Questionnaire.....	67
	Appendix C:	
	Administrators Questionnaire.....	71
	Appendix D:	
	OAT Study Sheet.....	72

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
1 Comparison of the Speaking Task Types.....	41
2 DBE B-level Teachers' and Testers' Experience.....	42
3 The Frequency of Subject Teachers' and Testers' OAT Participation.....	43
4 Comparison of the Speaking Task Types.....	50

CHAPTER 1 INTRODUCTION

Introduction

As is known, the English language is widely used outside English-speaking countries for business and diplomacy. Hence, large sums of money are invested in English language and learning. Consequently, we cannot escape from public interrogation about how efficiently and to what extent the money has been used, as related parties and individuals who made investments in this field want to be sure their money was wisely spent. As a result, language testing as one means of indicating the value of education becomes more and more important. The more language teaching and learning gains importance, the more related parties are interested in testing.

In spite of the importance of the testing, it is often regarded by educators to be a problematic issue in the language teaching and learning process as many teachers place some mistrust on tests themselves and testing as a procedure. Hughes (1990) elaborates the matter:

Mistrust is frequently well-founded. It cannot be denied that a great deal of language testing is of very poor quality. Too often language tests have a harmful effect on teaching and learning; and too often they fail to measure accurately whatever it is intended to measure. (p.1)

According to Hoekje and Linnel (1994), there is no certain consensus reached in second language assessment about an appropriate instrument to evaluate spoken language. In this respect, how the test content is determined, what procedure is followed and how it is scored are all critical considerations which continue to be

discussed.

Reliability, validity and practicality are three main measurement qualities with regard to testing. According to Brown (1996), reliability is attributed to a test when it gives the same results every time it is measured, and it measures exactly what it is supposed to measure (not something else). With respect to validity, he defines test validity as the degree to which a test measures what it claims, or purports, to be measuring. Although test reliability and validity are related to each other, they are different characteristics. Practicality is the third main issue in language testing and it is usually associated with the physical conditions under which a test is implemented. Time and money are the two features which are related to practicality. In respect to practicality in testing, Hughes (1990) says “other things being equal, it is good that a test should be easy and cheap to construct, administer, score and interpret”(p.47).

A related issue to validity is backwash effect. Normally an achievement test should measure what students have learnt in accordance with the course objectives specified in the syllabus of the course content. For example, if the teaching is comprehensive and appropriate, but testing is limited to few tasks, there may be harmful backwash effect. Hughes (1990) clarifies backwash effect as follows:

The effect of testing on teaching and learning is known as backwash.

Backwash can be harmful or beneficial. If a test is regarded as important, then preparation for it can come to dominate all teaching and learning activities.

And if the test content and testing techniques are at variance with the objectives of the course, then there is likely to be harmful backwash effect. (

p.1)

He also points out the relationship between teaching and testing as saying “There may be occasions when teaching is good and appropriate and the testing is not; we are then likely to suffer from harmful backwash” (p.2).

In this study, I focus on oral tests. My main concern is the content validity and backwash effect of the end-of-term Oral Assessment Test administered at the Department of Basic English of Hacettepe University. The Oral Assessment Test conducted at the Department of Basic English (DBE) is regarded as a final achievement test since it is administered at the end of the academic term. The main aim of this test is to measure what students have learnt during one term.

In general terms, oral assessment tests assess students’ communicative ability in English by asking them to respond orally under timed conditions to a variety of printed and aural stimuli that are designed to elicit both controlled and spontaneous responses. Since the Oral Assessment Test (OAT) in my study is an achievement test conducted to measure students’ oral language ability, consistency between the content of the oral test and learning goals set for the students and taught in the language program is of great importance in terms of content validity and backwash effect. What is normally expected is consistency between the syllabus, in which course objectives are set, and the content of the oral assessment tests. Henning (1987) points out that the content of a test should reflect the extent to which students have mastered the content of instruction. He also indicates that content validity is concerned with whether or not the content of the test is sufficiently representative and comprehensive for the test to be a valid measure of what it is supposed to measure. With respect to content validity, Bachman (1990) says that the

examination of content relevance and content coverage is a necessary part of the validation process since the domain specification upon which a test is based provides the means for examining it.

Background of the Study

The idea behind this study originated from overhearing complaints from teachers and students that the end-of-term OAT does not test what students have learnt within the related academic term. For this reason, I have launched this study to investigate the content validity and backwash effect of the OAT.

The Department of Basic English (DBE) is the preparatory school of English at Hacettepe University. The primary function of the DBE is to provide English language instruction mainly to Turkish students who are planning to continue studying at their departments and faculties where English is the medium in their disciplines, which include such diverse branches of science as Business Administration, Economics, Medicine, and some branches of engineering, such as Geology, Nuclear Energy and Food Engineering. The DBE offers courses in four skills: listening, writing, reading and speaking. The DBE has an overall curriculum which specifies general course objectives, but there is not a separate syllabus apart from that prescribed by the course books. The course books are chosen in accordance with the overall curriculum which frames and directs the instruction. For this reason, teachers follow the course content and abide by the course book objectives.

English language courses at the DBE are conducted at three levels, termed A, B and C levels. A level students consist of false beginners, those who received a borderline fail (50-59) on the placement test administered at the beginning of

September each year before the academic term starts. B level students are also false beginners, but consist of those who received low marks, ranging from 20 to 49 on the placement test. C level students are those who either got very low marks, lower than 19 on the placement test or those who have not taken the placement test. Though all A, B and C level students have to take an oral assessment test in addition to a written final achievement test administered at the end of each term, I have decided to focus only on B level students with the expectation that they would provide me more variation in the problem area as they consist of neither high nor low level of students.

When we look at the content of the course book, we can see that students are asked to perform a great variety of speaking tasks throughout the term. Some of them are independent tasks, but others are integrated into the skills such as reading and listening. Although they do many types of speaking tasks during the academic term, the content of the end-of-term OAT is limited to few speaking task types. In addition, in line with these claims, the OAT might be producing harmful backwash effect.

Statement of the Problem

English language instructors at DBE at Hacettepe University doubt whether the end-of-term Oral Assessment Test (OAT) tests students' overall oral language performance in accordance with what the course content offers in terms of the speaking tasks. Hence, the question of consistency between what students have learned with respect to oral skills in class what they have gained with the help of the speaking tasks within one academic term and what is tested in the end-of-term OAT arises. During the OAT, students answer a few personal questions as a warm up and

then talk on a topic drawn randomly by students. In the class, however, students practice several oral tasks in multiple ways, ranging from structured speaking activities such as role plays, discussions, debates and interviews to unstructured impromptu speeches and fluency practice in the framework of the course content within one academic term.

The questions about inconsistencies have led me to investigate the OAT at Hacettepe University to determine if the complaints are valid, and if so, in what ways.

Purpose of the Study

The aim of the study is to investigate the validity of the OAT and its backwash effects. In addition, I will take the subjects' perceptions and suggestions into consideration in terms of paving the way to improve the content of the exam and its classroom implications.

Significance of the Study

Brown (1996) stresses the importance of validity by saying “ Validity is especially important for all the decisions that teachers regularly make about their students. Teachers certainly want to base their admissions, placement, achievement and diagnostic decisions on tests that are actually testing what they claim to test” (p.231).

Since there is doubt on the part of teachers and testers as to whether the OAT is a true assessment of students' speaking performance, the validity of this oral test seems to be questionable. For this reason, this study will mainly be beneficial for my institution, the DBE of Hacettepe University, which operates as a department of the

School of Foreign Languages. In addition, weaknesses and strengths of the end-of-term OAT will be revealed, which will be of benefit to the teachers, administrators and students.

Subject teachers will be made aware of what this study found. Thus, they will have a chance to make informal decisions on possible revisions of the OAT. The students at the DBE will also benefit from the findings and conclusions of the study because any possible changes are directly related to the classroom learning.

This study might also be useful for other universities which deal with similar issues concerning oral assessment tests.

Research Questions

This study will address the following research questions:

To what extent does the end-of-term OAT represent the speaking items specified in the course book content for the preparatory language students at Hacettepe University, Department of Basic English (DBE) ?

Does the end-of-term OAT affect classroom teaching and learning (backwash effect), and if so, how?

CHAPTER 2 LITERATURE REVIEW

Introduction

This chapter presents an overview of the literature dealing with language tests and testing, but mainly validity and backwash effect of oral language tests. There are eight main parts in this chapter. First, it focuses on the definition, purpose and function of language tests. The second part is related to the types of tests. The third part contains general information on validity in terms of definition, types and importance. The next part deals with some difficulties encountered while administering and evaluating speaking tests. The fifth part projects a summary of instruments used for testing speaking. The sixth part is about the people involved in an oral test in terms of the definition and function of them such as interlocutors, assessors and candidates. Part seven highlights the assessment of an oral test in general and deals with the importance and difficulties of assessing oral language tests. The content of the last part is comprised of a review of the backwash effect of a language test, which emphasizes the effect of a language test on teaching and learning.

Definition, Purpose and Function of Language Tests

In general, a test is an instrument to measure quality and quantity and testing is a way or process to measure what is intended to measure. Various interpretations have been given to “language tests.” Bachman and Palmer (1996) say “Language tests are a valuable tool for providing information that is relevant to several concerns in language teaching. They provide evidence of the results of learning and

instruction, and hence feedback on the effectiveness of the teaching program itself” (p.8).

Carey (1988, p.xv) interprets language testing as saying; “Testing is an integral part of the teaching and learning process, and it provides teachers with vital information.” Madsen’s (1983) insights about language tests are as follows: “Language tests foster learning by their diagnostic characteristics. They confirm what each person has mastered, and they point up the language items needing further attention”(p.4). Henning (1987) elaborates on language testing in an interesting way. He focuses on elements and features other than the ones which we usually use in describing language tests and testing.

Testing, including all forms of language testing, is one form of measurement. Just as we weigh potatoes, examine the length of a piece of cloth, count eggs in a carton, or check the volume of a container of milk, so we test reading comprehension or spelling to determine to what degree these abilities are present in the learner. There is a potential for error when we weigh potatoes. For example, the scale might not work properly, or it may not be highly sensitive, so that we must settle for a rough estimate of the correct weight. Furthermore, the potatoes might be wet or dirty, or there might be a few yams mixed in. In either case our measurement may be inaccurate. (p. 1)

We language teachers are all aware that language testing is much more complicated than weighing potatoes as many other qualities are involved in language teaching. Students themselves, for instance, are much more complicated than potatoes. But it is useful to be reminded that there can be error in testing.

It is an undeniable fact that testing is an important part of every teaching and learning experience. A rough way to categorize a test is to label it as good or bad although we are aware of the fact that labeling depends on many other unconditional and changing situations. Good tests can sustain or enhance class morale and aid learning. Madsen (1983) points out the importance of good tests as stating “Good English tests also help students learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve” (p. 5). In connection with the quality of language tests, Madsen (1983) says “Properly made language tests create positive attitudes towards instruction by giving students a sense of accomplishment. Naturally, good language tests provide a better awareness of course objectives and personal language needs can help your students adjust their personal goals” (p. 4).

Bachman and Palmer (1996) expressed their views in a more detailed way when they emphasized the function and importance of language testing. They highlight the importance of good tests especially from the point of view of feedback on effectiveness of the teaching program and materials.

Language tests can also be a valuable tool for providing information that is relevant to several concerns in language teaching. They can provide evidence of the results of learning and instruction, and hence feedback on the effectiveness of the teaching program itself. They can also provide information that is relevant to making decisions about individuals, such as determining what specific kinds of learning materials and activities should be provided to students, based on diagnosis of their strengths and weaknesses, deciding

whether individual students or an entire class are ready to move on to another unit of instruction, and assigning grades on the basis of students' achievement. Finally, testing can also be used as a tool for clarifying instructional objectives and, in some cases, for evaluating the relevance of these objectives and the instructional materials and activities based on them to the language use needs of students following the program of instruction. (p.8)

In addition to focusing on the positive functions of tests, we need to remember the negative aspects too. Bad tests can ruin the whole system including the instruction in the classroom and affect negatively all the parties included.

Types of Tests

With respect to the test types, tests can be categorized into two main groups as norm-referenced and criterion-referenced tests.

Norm-referenced Tests

According to the definitions made by Brown (1996), norm-referenced tests are categorized into two types: proficiency and placement tests. He elaborates on norm-referenced tests saying, "norm-referenced tests are the ones that are used to compare the performances of students to each other" (p.v). This definition applies to both proficiency and placement tests. Brown also says

Norm referenced tests are commonly used to spread students out along a continuum of scores based on some general knowledge or skill area so that students can be placed, or grouped, into ability levels. The main purpose of these tests is to make comparisons in performance either between students within an institution, or between students across courses or institutions. Since

norm-referenced tests are to group students of similar ability, they mainly help administrators rather than teachers. (p.v)

Criterion-referenced Tests

According to Brown (1996) criterion-referenced tests can be categorized into two types: diagnostic and achievement tests. He says that criterion-referenced tests help teachers as they are administered to assess how much of the course material or sets of skills are taught in a course and learnt by the students. Brown (1996, p. vi) says “the purpose of the criterion tests is not to compare the performances of students to each other but, rather to look at the performance of each individual student vis-a-vis the material or curriculum at hand.” According to Brown (1996), these tests are usually used to diagnose the strengths and weaknesses of students with regard to the goals and objectives of a course or program. In other words, criterion-referenced tests are used to assess achievement, in the sense of how much each student has learnt. Such tests are useful to grade students’ performance in a course. These tests also help us improve the materials used, and sequencing of teaching points in a language program.

Comparison Between Proficiency and Achievement Tests

In Encyclopedic Dictionary of Applied Linguistics, edited by Johnson and Johnson (1998), proficiency and achievement tests are defined as follows:
 “Proficiency tests assess a learner’s level of language in relation to some absolute scale, or to the specifications of some job which has a language requirement.

Achievement tests assess how successful a learner has been in a course of study” (p. 187). As far as purpose and politics are concerned, it is said that proficiency tests are used as gate keepers. This means that proficiency tests provide opportunities for some by giving “a ticket” for access to a desired entity, or a refusal, which closes the gate in order not to admit them. In such ways, achievement tests are bound up with educational appraisal and management.

Hughes (1990, p.9) discusses the difference between proficiency and achievement tests as saying

Proficiency tests are designed to measure candidates’ ability in a language regardless of any training they may have had in that language. The content of a proficiency test is not necessarily based on the content or objectives of language courses. Rather, it is based on a specification of what candidates have to be able to do in the language in order to be considered proficient.

Concerning the function of proficiency tests, Hughes (1990, p.10) says “the function of these tests is to show whether candidates have reached a certain standard with respect to certain abilities”. Brown (1996) defines the proficiency test as a “gate opener” to an institution. He says “A proficiency test assesses the general knowledge or skills commonly required or prerequisite to entry into (or be exempted from) a group of similar institutions. One example is the Test of English as a Foreign Language (TOEFL), which is used by many American universities that have English language proficiency prerequisites in common” (p.10). It is clear that proficiency tests cannot be related to the goals and objectives of any particular language program. However, achievement tests are directly related to language courses and

their goals. Hughes (1990) categorizes achievement tests into two, as final achievement tests and progress achievement tests. The content of both final and progress achievement tests are based directly on a course syllabus or on the books and other materials used. The difference between them is that the former is administered at the end of a course or program, whereas the latter is conducted during the term or year. Midterm exams and pop quizzes are the examples of progress achievement tests.

Validity of Language Tests

Validity of language tests is defined in Encyclopedic Dictionary of Applied Linguistics (1998) as follows:

The validity of language tests, and in general of any measuring instrument like a performance sample, a questionnaire or an interview, is the extent to which the result truly represent the quality being measured. Traditionally, validity of language tests is estimated by internal criteria or content validity; comparison with other language tests or concurrent validity; comparison with other kinds of performance (such as occupation or subject examination) or predictive validity, or comparison with a theory of the performance in question (i.e. reading or listening comprehension, oral skills, or writing skill) or construct validity. (p.363)

With respect to validity in language tests, Henning (1987) projects his views in terms of purpose for which the test serves. Henning says that “Any test may be valid for some purposes, but not for others” (p.170). He supports his views as saying:

Validity in general refers to the appropriateness of a given test or any of its

component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what it is supposed to measure. (p.89)

Anderson, Clapham and Wall (1995) also stress the importance of purpose for which a test is designed. "The centrality of the purpose for which the test is being devised or used cannot be understated" (p.170). Anderson et al. (1995) expresses his views about validity from the point of purpose of a test as saying:

One of the commonest problems in test use is test misuse: using a test for a purpose for which it was not intended and for which, therefore its validity is unknown. This is not to say that a test cannot be valid for more than one purpose. However, if it is to be used for any purpose, the validity of use for that purpose needs to be established and demonstrated. (p.170)

Kitao and Kitao (1998) give another example "If the test purpose is to test ability to communicate in English, then it is valid if it does actually test ability to communicate. If it actually tests knowledge of grammar, then it is not valid test for testing ability to communicate. (p.1)

To sum up, the definitions made above have two important aspects. The first, validity is a matter of degree, which means that rather than saying the test is valid or not, it would be wise to deal with the degrees of validity, as some tests are more valid than others. Second, tests are valid or invalid in terms of their intended use.

There are four commonly discussed types of validity: content, criterion-related, construct, and face. (See Alderson et al., 1995; Bachman, 1991; Brown, 1996;

Heaton, 1988; Henning, 1987; Hughes, 1990;)

For the purposes of this study I will focus only on content validity.

Hughes (1990) indicates that “a test is said to have content validity if its content constitutes a representative sample of the language skills, structures etc., with which it is meant to be concerned. The test would have content validity only if it included a proper sample of the relevant structure” (p.22). Hughes stresses the importance of content validity as saying:

First, the greater a test’s content validity, the more likely it is to be an accurate measure of what it is supposed to measure. A test in which major areas identified in the specification are under-represented- or not represented at all- is unlikely to be accurate. Secondly, such a test is likely to have a harmful backwash effect. Areas which are not tested are likely to become areas ignored in teaching and learning. (p.23)

Heaton (1988) emphasizes the need for a careful analysis of the language test as far as content validity is concerned. He states that “the test should be so constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent” (p.160).

Difficulties of Testing Speaking

In spite of the variety of instruments developed and used to test speaking, many teachers feel less secure when dealing with tests which measure speaking ability than they do with standard pencil-and-paper tests. Madsen (1983, p.147) states that “the testing of speaking is widely regarded as the most challenging of all language exams to prepare, administer, and score. For this reason many people do not even try to

measure the speaking skill.” The reason for which testing oral ability is difficult may stem from the vague definition of what the nature of the speaking skill itself is. This clearly affects the process of determining validity.

One difficulty could arise out of choice of instruments used in testing speaking. Hoekje and Linnel (1994) say “No consensus has been reached in second language assessment about the appropriate instruments to evaluate the spoken language proficiency” (p.103).

“What is tested?” is another difficulty teachers and testers encounter. Taeduck and Finch (1998) point out that oral tests must be a true assessment of spoken abilities, rather than an indication of how well a student can produce well-memorized responses, and they claim that the issue of oral testing still highlights a major problem for educators. Many authors have tried to find positive and productive answers to the question of why speaking is the most difficult component in terms of language testing. Kitao and Kitao (1998) deal with one of the problems of testing speaking, pointing out the involvement of listening as a skill in speaking.

Success in speaking depends, to a great extent, on the listener. That is probably the reason why testing speaking does not lend itself well to objective testing. There are still questions about the criteria in terms of weighing for testing oral ability. It is difficult to separate the listening skill from the speaking skill. There is an interchange between listening and speaking, and speaking appropriately depends, in part, on comprehending spoken input. (p.1)

Another issue is that oral tests are the only tests in which testers are face to face

with candidates, which means that testers are in direct interaction with candidates.

What passes between testers and candidates is of primary importance. Therefore, the success in an oral test is not only related to the candidates' performance, but also the attitude and stand the testers take. Underhill (1987) emphasizes the matter saying:

In practice, success depends very much on the ability of the interviewer to create the right atmosphere, and it is a question of human personality.

It is a challenge to the interviewer to create the right atmosphere in a very short time, just as it is a challenge to the learner to respond to it. (p.45)

Underhill (1987) also stresses that the individual differences of the students in terms of personality are also of importance and should be taken into consideration.

He supports his views by saying:

Taking the initiative, asking questions, expressing disagreement, all require a command of particular language features. They also require the kind of personality. The natural instinct of many of us is to keep quiet, speak only when spoken to. There is therefore a danger that a discussion/conversation technique will reward extrovert and talkative personalities. (p.46)

The construction of the speaking test itself can also be problematic. Kitao and Kitao (1998) indicate that in some cases students are given a particular situation and instructed to respond in a certain way. In that case, students feel confined, as these tests are usually highly structured and require only a limited response, not connected discourse. They also stress the number of students as saying "testing speaking is also a particular problem when it is necessary to test large numbers of students, and even if each student speaks for only a few minutes, this becomes a huge job" (p.1).

In summary, we can say that there is still no consensus on how to measure speaking. Although many ways and instruments have been developed, one problem is defining the skill of speaking. In addition, personality features of both testers and candidates are also factors that affect the candidates' overall success and performance in an oral test.

The Instruments Used for Testing Speaking

A variety of instruments may be involved in testing speaking. With respect to the prompts given to candidates so as to make them speak, several means can be assigned. Using visual, verbal and written prompts are widespread in generating a conversation or a discussion. Various types of visual material might be appropriate for testing oral skills, depending on the language skill that the tester wants to elicit. For example, the official guide of First Certificate in English, published by UCLES in 1995 makes clear that "candidates are to be given visual prompts, such as photographs, line drawings, maps or diagrams in actual FCE exams. These visual prompts generate a discussion through engaging test takers in tasks such as planning, problem solving, decision making, prioritising or speculating" (p. 16). In addition to fluency, through careful selection of the material, the tester can control the vocabulary and grammatical structures required.

Role plays, discussions, oral interviews, simulations, descriptions and expressing opinions are all means of eliciting speaking. These are discussed below.

Role Play

In a role play the candidates are given information on which the role play is

based, and the candidates are evaluated on their ability to carry out the task in the role play. In the Longman Dictionary of Applied Linguistics, (Richards, J., Platt, J., & Weber, H. 1989), role play is described as follows:

Drama-like activities in which students take the roles of different participants in a situation and act out what might typically happen in that situation. For example, to practice how to express complaints and apologies in a foreign language. (p.246)

Discussion

Underhill (1987) defines discussion as “two people having a conversation on a topic of common interest” (p.45). In an oral test, discussion on a topic may take place between interlocutor and candidate or between two candidates. Both parties exchange their opinions on a common topic. In order to make the definition for discussion as a task clear, Underhill states that “The task usually involves taking information from written documents and coming to a decision or consensus about the topic through the discussion” (p.49).

Oral Interview

Oral interviews are usually testing situations in which the tester generally has a list of questions to ask the candidate, and someone, either the interviewer or another person but preferably another person- assesses the language proficiency of the candidate.

Underhill (1987 p.54) says “ The interview is the most common of all oral tests. It is a direct, face to face exchange between learner and interviewer. It follows a pre-determined structure, but still allows both people a degree of freedom to say what

they genuinely think.”

Simulation

Richards et. al. (1989) in the Longman Dictionary of Applied Linguistics defines “simulation” as follows:

Simulations are activities which reproduce or simulate real situations and which often involve dramatization and group discussion (Role Play does not include group discussion). In simulation activities, learners are given roles in a situation, tasks or a problem to be solved, and are given instructions to follow (for example, an employer-employee discussion over wage increases in a factory). The participants then make decisions and proposals. (p. 259)

Description

In oral tests, there might be tasks based on some kinds of descriptions, such as description of people, objects, or and event or procedure. Some possible topics which are asked in the OAT at Hacettepe University and others that candidates could be asked to talk about in an oral test (See eg. Underhill, 1987, p.69) may be as follows:

- Describe the stereotypical Turkish man or woman.
- What is the definition of a successful person for you?
- Can you describe someone you like/admire or dislike very much?
- Describe how people in your country celebrate the New Year.
- Give instructions for using a public pay-phone.
- Describe a bicycle.

With respect to the tasks based on description, Underhill (1987) says that, “The choice of topics can make the tasks more or less controlled. A question such as “Describe your favourite meal” would be less controlled as there can be a lot of possible answers; whereas “Explain how you change a car tyre” has basically only a simple answer” (p.70).

Expressing Ideas and Opinions on a Topic

In an oral test, candidates may be asked to respond to a discursive topic which may either be chosen by the candidate or be presented to him or her by the testers. Through the topic or topics, candidates are expected to express their own opinions to justify, support or simply explain the reasons for or against the topic.

Underhill (1987) clarifies a task which is based on eliciting candidates’ opinions on a topic as follows:

Candidates are invited to choose a discursive topic to speak on at a few minutes’ notice. These would usually be topics of current interest on which everybody who follows current affairs is pressured to hold an opinion. As well as explaining his own position on his chosen issue, the candidate is invited to give reasons supporting his position; and when he has finished speaking, the interlocutor may ask questions to clarify a point or to explore further the arguments presented. (p.70)

Some sample topics (See Underhill, 1987, p.70) could be as follows:

- Do you favor the increasing use of nuclear energy? Why?
- What would be your first act as Prime Minister?
- Why do women do so much work and receive so little money?

The choice of the right topic or topics is also an important issue as they are used to encourage candidates to talk on and to measure their overall speaking performance. This especially for the instruments “Discussion”, “Description” and “Expressing Ideas and Opinions on a Topic”. In some oral tests, topic elicitation is the only means to measure oral language performance of students. For example, in the DBE of Hacettepe University, the end-of-term OAT is based solely on topic elicitation. Underhill (1987) emphasizes the importance of choosing topics to generate a conversation or discussion, saying:

Choosing the topic is very important. It should be relevant to the aims of the program or the needs of the learners and should contain new information or put over a new point of view. It should not be so specialized that only the speaker himself is interested, nor should it be so general that it has no apparent purpose other than as a language exercise. Ideally, the topic should be chosen by the learner in consultation with his teacher who will help match the ability of the learner with the difficulty of a given topic. Some learners will play safe by choosing the topic they are most familiar with. However well prepared a speaker is, he will not be able to talk as confidently about a new topic as he will about the one he already knows well. The assessor has to be careful to take this into consideration: is the topic in itself a difficult one irrespective of the fluency of the speaker. (p.47)

To sum up, there are various speaking tasks that are used to measure candidates' speaking performance. The broad aim of all these tasks is to encourage candidates to

speak by giving them prompts and stimuli to speak about. What should always be taken into consideration in terms of the speaking tasks in an oral test is that the purpose and the content and also the quality of the speaking tasks are of great importance. Weir (1990) says “.....speaking tasks developed should be purposive, interesting and motivating, with a positive washback effect on teaching that precedes the tests” (p.73).

Related Parties Involved in an Oral Test

In a large testing program, such as at Hacettepe University, different duties are carried out by different educational members with different qualifications and skills.

- Testers to develop the tests
- Test writers, proof readers and also those who print and pack the tests
- Testers to administer the tests as interlocutors or interviewers
- Testers assigned to assess as assessors or markers to mark or remark the tests afterwards
- Administrators to carry out administrative duties

Tester, as used above, is a general term for a person who is in charge of test conducting. Equally, a person who either prepares or administers a test can also be called “tester”. Testers are called in some different ways according to what role they undertake in an oral test; for example, as interviewers, assessors and markers.

The definitions below were taken from the book “Testing Spoken Language” written by Underhill (1987)

- **Interlocutor:** Some oral tests have a person whose job is solely to help the candidate to speak, but who is not required to assess him. An

interlocutor is a person who talks with a learner in an oral test , and whose specific aim is to encourage the candidate to display, to the assessor, his oral fluency in the best way possible. An interlocutor is not an assessor.(p.7)

- **Assessor:** An assessor is a person who listens to a candidate speaking in an oral test and makes an evaluative judgement on what she hears. The assessor will be aided by pre-defined guidelines such as rating scales, which give considerable help in making these judgements. (p.7)
- **Marker:** This term is reserved for someone who is not present at the test itself but later awards marks to the candidate on the basis of an audio or video tape recording. (p.7)
- **Interviewer:** An interviewer is a person who talks to a candidate in an oral test and controls to a greater or lesser extent the direction and topic of the conversation. An interviewer may also take the role of assessor or one of the assessors. (p.7)

The word “candidate” is general term for a person who takes a language test of any kind. This person might be a student in an achievement test, or a person who is not a student but a only a test taker or testee in a proficiency test. For example, a test taker could also be an applicant for a job or a program in which oral language proficiency is required. A candidate can also be called as an interviewee in an interview.

Assessment Criteria for an Oral Test

Assessment is probably the most critical issue in oral tests. What is tested in oral tests in terms of assessment criteria may differ according to the type or aim of the

oral tests. The assessment criteria for oral proficiency tests may be different from the ones used for oral achievement tests. Moreover, the level of the students necessitates different criteria.

With reference to the speaking part, Paper 5 of the Cambridge First Certificate in English exam (FCE), which is an upper-intermediate level of exam, Haines and Steward (1997, p. 16) summarize the criteria for assessment as follows:

During the test each candidate is assessed according to the following criteria:

- use of grammar and vocabulary
- pronunciation
- ability to communicate effectively (interactive communication)
- fluency

In addition to the FCE criteria given above, in International English Language Testing System, known as IELTS, Jakeman and Mc Dowell (1996) state that “the assessment criteria involves ‘the ability to ask questions’ in which candidates must ask the examiner questions in response to the given cue card that describes a situation or problem” (p. 7).

The criteria used to assess the advanced level of Paper 5 of the Cambridge Proficiency in English examination is clarified by Gude and Duckworth (1997, p. v) as follows:

Fluency : Speed and rhythm, choice of structures, general naturalness and clarity.

Accuracy : Control of structures including tenses, prepositions, etc. to an effective level of communication.

Pronunciation (Individual sounds) : Correct use of consonants and vowels in stressed and unstressed position for ease of understanding.

Pronunciation (Sentences) : Stress timing, rhythm and intonation patterns, linking of phrases.

Interactive communication : Flexibility and linguistic resource in exchange of information and social interaction.

Vocabulary : Variety and correctness of vocabulary in the communicative context.

Backwash Effect

“The effect of testing on teaching and learning is known as backwash. Backwash can be harmful or beneficial” (Hughes, 1990, p.1).

The term “backwash” is interchangeable with “washback”. It is up to the author’s preference to use either “backwash” or “washback” as the effect of testing on teaching and learning is considered. Anderson and Wall (1993) make the two different terms clear. “This phenomenon is referred to as “backwash” in general education circles, but it has become to be known as “washback” in British applied linguistics” (p.115).

Frederiksen and Collins (1989, cited in Anderson and Wall,1993, p.116) discuss the notion of washback validity by using the term as “systematic validity.”

A systemically valid test is one that induces in the education system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure. Evidence for systemic validity would be an improvement in those skills after the test

has been in place within the educational system for a period of time.

(Frederiksen and Collins. 1989, p.27)

Anderson and Wall (1993) explore the notion of washback and a series of possible Washback Hypotheses. There were 15 hypothesis listed. As for example, three of them were given below.

- A test will influence teaching.
- A test will influence learning.
- A test will influence attitudes to the content, method, etc. of teaching and learning. (p.120-121)

They pointed out that “the notion of washback is common in the language teaching and testing literature, and tests are held to be powerful determiners of what happens in classrooms” (p.115). They also suggest that a test’s validity should be measured by the degree to which it has beneficial or harmful influence on teaching. They assert that rather than saying beneficial or harmful backwash effect, it could be wise to approach the matter by seeking to find an answer to the question inquiring to what extent a tests affects the teaching and learning in the classroom.

The issue of washback effect of a test on teaching and learning seems to be simple, but in fact it is quite complex. Anderson and Wall (1993) imply that there might be other forces which are involved in the nature of washback effect. “It is not at all clear that if a test does not have the desired washback this is necessarily due to a lack of validity of the test” (p.116). They claim that other forces within society, education and schools might prevent washback from appearing. These forces can hardly be attributed to a problem with only the test itself. That is the reason why

they assert that validity is a property of a test whereas washback is likely to be a complex phenomenon which cannot be directly related to a test's validity.

Bachman and Palmer (1996) gather what previously was revealed in terms of washback and refined it through their views summarizing as follows:

Washback has been discussed in language testing largely as the direct impact of testing on individuals, and it is widely assumed to exist. However, washback has potential for affecting not only individuals but also the educational system as well, which implies that language testers need to investigate this aspect of washback also. (p.30-31)

Bachman and Palmer (1996) finalize their views as saying "Thus, in investigating washback one must be prepared to find that it is far more complex and thorny than simply the effect of testing on teaching" (p.31).

This review of the literature on validity and backwash effects of language tests focuses on the complexity of both issues. This complexity leads to the necessity of further investigation into validity and backwash effects of the Oral Assessment Tests administered at Hacettepe University. In the next chapter, I will describe my methodological approach to this investigation.

CHAPTER 3: METHODOLOGY

Introduction

The main concern of this study is to investigate the content validity and backwash effect of the end-of-term Oral Assessment Tests (OAT) administered at Hacettepe University, Department of Basic English (DBE). The Oral Assessment Tests conducted at DBE are administered at the end of each academic term, both fall and spring terms as an achievement test with the main aim being to measure what students have learnt in terms of oral skills during one term.

In my study, the consistency between the content of the OAT and program's learning objectives set for the students and realized within the frame of the language program is my main concern; hence, the primary research question is "To what extent does the end-of-term OAT represent the speaking items specified in the course book content for the preparatory language students at Hacettepe University, Department of Basic English (DBE)?" Apart from the consistency between the course book content and the content of the OAT, I am interested in the backwash effect of the OAT, which is the effect of the OAT on teaching and learning. Therefore, this study covers the examination of the backwash effect in line with the following research question: "Does the end-of-term OAT effect classroom teaching and learning (backwash effect), and if so, how?"

This methodology section contains four sub-sections. The first section provides information on the informants used in the study. Second, the materials used in the study are explained. The third section provides the information on how the study was conducted. Finally, the data analysis section describes how the data were

arranged and analyzed.

Subjects

The DBE, which is the preparatory school of English at Hacettepe University, provides English language instruction mainly to Turkish students who are planning to continue studying at their departments and faculties in English.

English Language courses at the DBE are conducted at three levels: A, B and C, as described in Chapter 1. A-level students consist of false beginners as they receive a borderline fail (50-59) on the placement test administered at the beginning of September each year before the academic term officially starts. B-level students are also false beginners but they are those who receive lower marks ranging from 20 to 49 on the placement test. Those who get very low marks, 19 or lower, on the placement test, or those who did not sit for the placement test make up the C-level students. For this study, the subject chosen was B level. The informants in this study are B-level subject teachers and testers, B-level students, and the administrators at DBE.

B-Level Subject Teachers and Testers

There were 14 B-level teachers teaching B-level classes and 2 testers at Hacettepe University (HÜ). As well as fulfilling their duties as testers, which included preparing written and oral tests, the testers were teaching B-level students 6 hours a week. Both the 14 subject teachers and the 2 testers were given questionnaires and they all responded. Of the 16 teachers and testers, 14 had been working for 1-5 years; 2 had a working experience ranging for 6-10 years at HÜ. At that time, two of them had been assigned to administer the OAT as assessor and

interlocutors more than 11

times; 8 of them 4-10 times and six of them 1-3 times. Their opinions were valued as important as they would help to give insights about the problem areas

B-Level Students

There are 348 B-level students at Beytepe Campus of Hacettepe University. I considered 20% of the B-level students to be an appropriate sample. As a result, 65 randomly selected students were given questionnaires just after the end-of-term OAT was administered on May 18, 1999. Out of the 65 students, who were from different classes, 62 of them responded to the questionnaire, for a response rate of 95.38 %. Of the 62 B-Level students, 39 of them were females; the rest were males. Their age range was 17-21.

When the end-of-term OAT is administered, the language level of the students is required to be upper-intermediate because the course book, New First Certificate Masterclass, is an upper intermediate level of book, and by the end of the academic term students who entered as false beginners reach the level of upper intermediate.

Administrators of the DBE

Three administrators, one of whom is an academic coordinator, were given questionnaires in order to ascertain their ideas and suggestions in terms of the end-of-term OAT.

Materials

For this research study, three types of materials were used. The first one was the course book, New First Certificate Masterclass, written by S. Haines and B. Steward and published by Oxford University in 1996. The course book was used to

identify the types of speaking tasks taught in B-level courses. The second kind of materials was the speaking topics (Appendix D), which had been given to B-level students as the OAT study sheet two weeks before they sat for the end-of-term OAT. The OAT study sheet consisted of 65 randomly selected topics from different sources including the course book. In the OAT all the students were held responsible for all 65 topics in the OAT study sheet.

The third type of materials used in this study was questionnaires. Three different types of questionnaires were developed (See Appendices A, B, and C), and given to B-level subject teachers and testers, B-level students, and administrators. The first type of questionnaire was given both to B-level teachers and testers. There are four sections in these questionnaires. The first section was about their teaching experience and the frequency of their previous OAT test participation. The second section was designed to reveal their awareness of the relationship between the test content and the course content. In the third section, there were six questions with the aim of revealing backwash effect of the OAT on the language teaching in the classroom. In section 4, there were three open-ended questions asked to elicit the subject teachers' and testers' attitudes and opinions about the OAT.

The second type of questionnaire was for the B-level students. There were 8 questions in two sections. The questions consisted of either "YES or NO" or "multiple choice type of questions" giving some variety of options to mark.

The third type of questionnaire was for administrators. In this questionnaire there were only three questions, the first of which was yes-no question. The other two questions were linked to administrators' response to the first question, the content of

which was about their evaluation of the OAT. The second question was for those who responded as “YES”, the third was for those who responded “NO”. The aim of this questionnaire was to obtain their attitudes towards the OAT.

All three types of questionnaires were piloted beforehand to allow for revision as a result of any difficulties in understanding the items of the questionnaires. In order to pilot students’ questionnaires, two students were asked to fill out the questionnaires. For teachers and testers’ questionnaires, two experienced teachers were employed to pilot them. For the administrators’ questionnaires, one of the veteran administrators was asked to fill it out. Following this, I revised the questionnaires in order to clarify the questions.

Procedures

The questionnaires were handed out to B-level subject teachers and testers, B-level students and administrators at Hacettepe University DBE respectively. Before distributing questionnaires, the participants were informed about the purpose of the research.

First, all 14 B-level subject teachers and 2 testers were given questionnaires and they all responded to the questionnaires for a response rate of 100%. The questionnaires were about the evaluation of the content of both the end-of-term OAT and the course syllabus. In these questionnaires the subject teachers and testers were also required to give their views about the backwash effect and possible ways of improving the OAT. The B-level subject teachers and testers were given one week for the completion of the questionnaires after distribution.

Second, 65 B-level students were given questionnaires and 62 of them responded

to the questionnaires for the response rate of 95.38 %. The questionnaires had been designed and worded to be answered quickly since they were given out just as students emerged from the OAT on May 18,1999. Students were not timed specifically. Upon finishing filling them out, they handed in the questionnaire forms. The aim was to receive their fresh and vivid views of the OAT itself, about the content of the OAT and the preparation procedure they supposedly had done for that exam.

Third, administrators' questionnaires were conducted one week after the end-of-term OAT was administered. The three administrators were given one-page questionnaire forms, and three days later they were collected for a response rate of 100 %.

Data Analysis

To analyze the consistency between the types of speaking tasks in the course book and those the students were held responsible in the end-of-term OAT, first of all, I examined the speaking tasks in the course book content and then identified the types of speaking tasks in each unit. Following this, I computed the frequency of the speaking task types in all 14 units. The frequency was calculated as follows: first of all, the types of speaking tasks in one individual unit was found. Then, the same procedure was applied to the other units. Finally, the occurrence of each speaking task type across 14 units was calculated. For example, "role play interview" existed in only one unit, whereas "picture discussion" took place in ten units.

The second step for analysis procedure was to analyze the 65 speaking topics which were given to all B-level students as a study sheet two weeks before they sat

for the end-of-term OAT. I categorized these 65 topics into speaking task types, such as expressing ideas and opinions, description and narrative. I then calculated the number of separate types and found the frequencies. Categorization for the speaking tasks in the OAT was based on the criteria explained in Chapter 2, under the sub-heading of Instruments Used for Testing Speaking.

The aim of the first and second steps of the data collection procedure was to be able to compare the speaking task types occurrence in the course book with those in the end-of-term OAT.

The speaking tasks in the course book were listed and counted to determine the frequencies. Then, the speaking tasks through which students were tested in the OAT were listed in types, and their frequencies were counted. Following this, the types of the speaking tasks in the course book were compared with the ones in the OAT with regard to their existence and frequencies.

The third step of data analysis was to analyze the questionnaires. The three types of questionnaires contained mixed question types. The data obtained from the yes-no questions and multiple choice questions were analyzed by frequencies and percentages. The answers to the open-ended questions were analyzed by putting them into categories according to “recurring themes. In addition, some striking points mentioned by the respondents were directly quoted.

This chapter has discussed the subjects included in the study, the materials used in the research design and the procedure and data analysis techniques used. In the next chapter the results of the data analysis are displayed and discussed.

CHAPTER 4: DATA ANALYSIS

Overview of the Study

This study investigated the content validity and backwash effect of the end-of-term Oral Assessment Test (OAT) administered at Hacettepe University, Department of Basic English (DBE).

To collect data, first of all, upon examining the speaking parts of the course book, I identified the types of the speaking tasks in the course book. Secondly, I examined the speaking tasks of the end-of-term OAT. Then, I compared the task types in the course book syllabus with the tasks types in the end-of-term OAT in order to determine the consistency between them. The aim was to reveal content validity of the OAT.

In addition, in order to investigate the content validity and backwash effect of the end-of-term OAT, three separate questionnaires (See Appendixes A,B and C) were given to three groups of subjects: 14 B-level subject teachers and two testers, 65 randomly selected B-level preparatory school students and 3 administrators. Their opinions were valued as important, as they would help to give insights about the problem areas. The aim of the questionnaires was, first of all, to learn their general attitude towards the OAT, and second, in accordance with the sequence of the questions, to get their opinions about the content of the OAT. In addition, through the questions of the questionnaires, I wanted to find out how the OAT affected the learning and teaching in the classroom in order to reveal the backwash effect of it. Data analysis procedure is described in the subsequent section.

Data Analysis Procedure

The data collected were analyzed using the following procedure. First, I documented the types of the speaking tasks in the course book content. To do this, I identified the types of the speaking tasks in each unit. Then, I computed the frequency of the each type of the speaking tasks in all 14 units. For documentary purposes, I am interested in the occurrence of the task types across 14 units, not the number of the speaking tasks existing in each unit. In addition to indicating the frequency of speaking task types in the course book syllabus, I examined the 65 topics (See Appendix D) which were given to all B-level students before they took the end-of-term OAT. I categorized these 65 topics in terms of task types and determined their frequencies. Finally, the types and frequencies of the speaking tasks in the course book syllabus were compared with those for which students were held responsible in the OAT.

In addition to the documentation of the speaking tasks in both the course content and the end of term OAT, for the purpose of collecting data, three types of questionnaires were developed and given to B-level subject teachers and testers, B-level students and the three administrators respectively.

The first type of questionnaire (See Appendix A) was given to 14 B-level subject teachers and two testers, all of whom responded. There were four sections in the questionnaires. The initial section was about their experience and frequency of previous OAT participation. The second section was designed to reveal their awareness of content of both the course book and the test. In the third section, there were multiple choice type of questions with the aim of obtaining their opinions about

content validity and backwash effect of the end-of-term OAT. To analyze the findings obtained from the questionnaires, frequencies and percentages of the responses were taken. The last section was comprised of open-ended questions, which were analyzed by putting responses into categories according to recurring themes.

The second type of questionnaire (See Appendix B) was given to B-level students just after they had taken the end-of-term OAT on May 18, 1999. Out of the 348 B-level students, 65 students (18.67%), who were from different classes, were randomly given questionnaires. Of those, 62 of them responded the questionnaire for a response rate of 95 %. There were 8 questions in two sections. The first section asked about biographical data, such as gender and age group. The second section was concerned with students' feedback on the end-of-term OAT as to whether they found the OAT easy or not and whether they used the course book for OAT preparation. In addition, there were questions about what type of speaking tasks were involved in the OAT. The questions were either yes-no questions or multiple choice types of questions. The responses were evaluated by frequencies and percentages.

The third type of questionnaire (See Appendix C) was given to the three administrators at DBE so as to ascertain their ideas and suggestions in terms of the content validity of the end-of-term OAT. In these questionnaires, there were only three questions, all of which were designed to elicit their views and opinions about the end-of-term OAT. Their responses were evaluated according to recurring themes.

Results of the Study

Analysis of the Speaking Tasks

In Table 1, the types of the speaking tasks in the course book and in the content of the end-of-term OAT are displayed in order to be able to compare the consistency between them.

Table 1

Comparison of the Speaking Task Types

Task Types in Course Book	Number of Units the Task Occurs in N=14		Frequencies of the Task Types in the OAT Study Sheet N=65	
	f	%	f	%
Role Play	4	(28.5%)	—	
Role Play Interview	1	(7.1%)	—	
Information Gap	4	(28.5%)	—	
Picture Description and Discussion	11	(78.5%)	—	
Description of Place, Object, People or Social Events	5	(35.7%)	6	(15%)
Expressing Ideas and Opinions	14	(100%)	58	(93.5%)
Discussion on a Discursive Topic	14	(100%)	—	
Debate	1	(7.1%)	—	
Problem Solving	6	(42.8%)	—	
Making Plans	3	(21.4%)	—	
Simulations	6	(42.8%)	—	
Making Decisions	5	35.7%	—	
Narrative	6	42.8%	1	(1.6%)

Table 1 indicates that there are 13 different types of speaking tasks in the course book content, ranging from role plays to narrative tasks. However, in the end-of-term OAT, there are only three types of speaking tasks. It is notable that the two

speaking task types, Picture Description and Discussion and Discussion on a Discursive Topic which exist in the course book with a high frequency are not included in the OAT study sheet.

Analysis of the Questionnaires

There were three groups of subjects involved in the study. They were B-level subject teachers and two testers, B-level students and the three administrators. The results of the questionnaires are reported below.

Subject Teachers and Testers Questionnaire. The questionnaire (See Appendix A) was given to all 14 B-level subject teachers and two testers, all of whom responded. There were four sections in the questionnaires.

The first section was about their experience and frequency of their previous OAT participation. The related figures are displayed in Tables 2 and 3.

Table 2

DBE B-level Teachers' and Testers' Experience (N=16)

	Overall Teaching			Teaching Experience		
	Experience			at DBE		
	1-5	6-10	11-15	1-5	6-10	11-15
Subject Teachers	5	5	4	13	1	-
Testers	1	1	-	1	1	-

Of the 14 subject teachers, nine teachers have 6 or more years of overall teaching experience, but almost all teachers do not have more than 5 years at the DBE at Hacettepe University. One of the testers' teaching experience at DBE ranges

between 6-10 years, which means she has got considerable teaching experience.

Both testers gained their teaching experience at the DBE.

Table 3

The Frequency of Subject Teachers' and Testers' OAT Participation (N=16)

	More than 11 times	4-10 times	1-3 times	Never
B-level Subject Teachers	2	6	6	
Testers		2		

In comparison of the experiences of the end-of-term OAT participation, it is clear that both testers have participated in the OAT at least four or more times.

Section 2 of the questionnaire was designed to reveal subject teachers' and testers' awareness of the task types in the course book and in the OAT. The participants were given the same list of the task types as the ones used in Table 1 and they were asked to put a tick next to the tasks which are assessed in the end-of-term OAT. According to the responses given to this section by the 16 participants, the speaking tasks, Description of Place, Object, People or Social Events received 15 ticks (93.7%), Expressing Ideas and Opinions 16 (100%), Discussion on a Discursive Topic 3 (18.75%), Problem Solving 2 (12.5%), Making Plans 1 (6.25%) and Making Decisions 2 (12.5%). The rest of the speaking task types were not ticked. (See Table 4, p. 49)

Section 3 of the questionnaire consisted of six multiple choice type of questions. Since the choices are different in all six questions, the results are individually evaluated.

In Question 1 the participants were asked whether the end-of-term OAT represented all the speaking task types specified in Section 2. Of the 16 respondents, 13 (81.25%) disagreed, and 3 (18.75%) strongly disagreed with the item.

Question 2 inquired how much the tasks which are assessed in the end-of-term OAT represent the whole course content. Of the 16 respondents, 11 of them (81.25%) said that the OAT was representative “to some extent”, and 5 respondents (31.25%) claimed that there was little representation.

Question 3 asked the 16 participants if the tasks which are included and assessed in the end-of-term OAT were adequate to measure students’ overall speaking performance. 15 respondents (93.75%) marked the choice “Not adequately, but to some extent”. The other respondent said that the tasks in the OAT did not measure students’ overall speaking performance.

Question 4 aimed to reveal the value of the speaking tasks that the respondents had not ticked in Section 2 because of the claim that they were not assessed in the OAT. The value of the task types in terms of students gaining and improving speaking skills was questioned. Of the 16 respondents, nine respondents (56.25%) marked the choice “very valuable” and seven respondents (43.75%) said that “valuable”.

With the aim of revealing the backwash effect of the OAT, Question 5 asked the subject teachers and testers if they thought students put a lot of effort and time into performing well in the classroom on those speaking tasks which would not be assessed and therefore, are of no value in terms of pass and fail degree in the OAT. Of the 16 respondents, only two of them (12.5%) said “Yes, they put a lot of effort”.

Ten respondents (62.5%) marked the choice of “Yes, but to some extent”. Four respondents (25%) said “Not very much”. Nobody marked the last choice “Not at all”.

Question 6 was for those who marked either “Yes, but to some extent” or “Not very much” in response to Question 5. 14 respondents out of 16 had marked these two options. The purpose of Question 6 was to learn the reason why the teachers and testers thought students put less effort and time into performing well in the classroom on those speaking tasks. Eleven of them agreed with the statement, “The reason is that students know these tasks are not assessed and are of no value in terms of pass or fail degree in the OAT. As a result, the OAT affects the learning and teaching in the classroom somewhat negatively”. However, three of them disagreed with this statement.

Section four consisted of three open ended questions which were analyzed by putting the responses into categories according to recurring themes.

Question 1 asked the 14 subject teachers and 2 testers whether they were pleased with the end-of-term OAT and why/why not. Of the 16 respondents, 2 (12%) were not pleased very much, 14 (87.5%) were not pleased with the OAT at all. The reasons for their dislike were as follows.

- The high number of students tested in one day.
- The time given each student was not longer than 5 or 6 minutes which was too short to show their speaking performance.
- The topics, some of which were too difficult, some too easy or restricted to talk on, do not appeal to students background, experience or interest.

- The content of the OAT was too narrow in terms of task types to represent the course content and to assess students' speaking performance in different ways.

Question 2 was about the possible ways to improve the OAT. The responses given by the 16 respondents can be categorized under the following suggestions:

- The number of students to be tested in one day should be decreased.
- Students should be given longer time in the OAT to express themselves and to show their speaking performance in a relaxed atmosphere.
- Delicate attention should be given to the quality and content, and selection of the topics.
- A greater variety in speaking task types should be provided. They recommended that discussions, making plans, making decisions, picture discussions, and information gap types of speaking tasks should be involved in the OAT.
- Oral assessment tests should not be given at the end of a term only. It must be administered at least twice a term. The respondents say that because there are no oral tests within the term, most of the speaking activities are ignored by both teachers and students in the classroom. They referred to backwash effect with this suggestion.

In question 3, subject teachers and testers were asked if they thought that speaking as a skill should be emphasized more in the classroom. In addition, they were asked to explain their opinions in accordance with their initial response of "Yes" or "No". Of the 16 respondents, 15 of them responded with "Yes"; however, one of them said "No". When I analyze the reasons behind "Yes" responses, 7 respondents out of 16

said that emphasis should be given to speaking equally to grammar and vocabulary. Six respondents stressed other skills, saying “speaking” as a skill should be emphasized as much as the other skills such as “reading” and “writing”.

B-level Students Questionnaire The following responses are from 62 B-level students. There were two sections in the questionnaire. (See Appendix B) The first section asked about biographical data, such as gender and age group. Of the 62 students, 29 (46.7%) were males, 33 (53.3%) were females. Fifty-four students (87%) were between the ages of 17-20, and 8 students (13%) were between 21-25 years old.

The second section of the questionnaire contained 6 questions with the aim of obtaining students’ reflections in terms of the content and backwash effect of the end-of-term OAT. The questions were either yes-no questions or multiple choice types of questions.

Question 1 asked the students whether they found the OAT easy. The answers are fairly equally divided. Of the 62 students, 15 (24.19%) said they had found the oral test quite easy, 24 (38.70%) somewhat easy, 23 (37.09%) not very easy.

Question 2 was asked to learn if students did any special study for the end-of-term OAT in their classrooms with their subject teachers. In response to this question, of the 62, 52 students (83.87%) said “YES”, 10 students (16.12%) “NO”.

Question 3 addressed those who responded to Question 2 as “YES”. It asked the students whether they only used outside material which their teacher brought into the classroom. Of the 62 students, 27 students (51.92%) said “YES” while the rest, 25 students (48.07%) said “NO”.

Question 4 was for those who responded to Question 3 as “NO”. Twenty-five students responded to Question 4, which asked whether they used the course book and outside materials equally, or mostly the course book, or mostly the outside materials for the OAT preparation in the classroom. Of the 25 respondents, 7 students (28%) said they used both of them equally, 1 student (4%) said they used mostly the course book and 17 students (68%) said they used mostly outside materials.

Question 5 inquires whether the course book contributed much to their oral exam preparation. Of the 62 students, 15 students (24.19%) marked the choice of “yes, a lot”; 22 (35.48%) “somewhat”; and 25 students’ choice (40.32%) was on the option of “not very much”. The aim of Questions 3, 4, and 5 was to reveal backwash effect in terms of the course book’s usage and contribution to the OAT preparation. The answers seem to indicate that the backwash effect of the OAT is quite far from a beneficial (positive) backwash effect, which is ideally expected from an oral assessment test.

Question 6 aimed to reveal students’ awareness related to the speaking task types in the end-of-term OAT. Accompanying this question was the same list of speaking task types as those used in Table 1. The students were asked to put a tick next to the task types on which they were assessed in the end-of-term OAT. The results were analyzed not only from the students’ point of view related to the speaking task types, but I also compared B-level students’ awareness with B-level subject teachers’ and testers’ awareness. Table 4 summarizes the responses.

It is interesting to know that although the two tasks, Discussion on a Discursive

Topic and Making Decisions do not exist in the OAT, of the 16 teachers and testers, three of them (15.75%) and of the 62 students, 11 of them ticked Discussion on a Discursive Topic; likewise, of the total teachers and testers, two of them (12.5%) and of the total students, nine of them (14.5%) ticked Making Decision, claiming that they occur in the OAT. The reason may have arisen out of their different interpretation of these types of speaking tasks. They may have assumed that although one student takes the test at a time and there is no discussion on a topic with another student, interlocutor' or assessors' involvement creates a discussion. However, as it is explained in Chapter 2, Discussions are defined as occurring between two students, or with the involvement of more than two students. Equally, Making Decision as a speaking task requires a situation in which two students discuss the matter and make their decisions.

Apart from those, when the results of teachers' and testers' opinions are compared with the students' opinions in terms of the speaking task types in the OAT, it can be seen that there is a more or less of a consistency between the teachers' and students' responses, which means there is a somewhat agreement between both groups of subjects in terms of their perceptions of the speaking task types occurring in the OAT.

Table 4

Comparisons of the Speaking Task Types

The Speaking Task Types In the Course Book N=13	Number of Units the Task Occurs in N=14	Frequencies of the Task Types in the OAT Study Sheet		Teachers' Perception of the Task Types in the OAT		Students' Perception of the Task Types in the OAT	
		N=65 f	%	N=16 f	%	N=62 f	%
Role Play	4	-	-	-	-	-	-
Role Play Interview	1	-	-	-	-	-	-
Information Gap	4	-	-	-	-	-	-
Picture Description and Discussion	11	-	-	-	-	2	3.2%
Description of place, object, people or social events	5	6	9.6%	15	93.7%	48	72.4%
Expressing Ideas and Opinions	14	58	93.5%	16	100%	62	100%
Discussion on a Discursive Topic	14	-	-	3	15.75%	11	17.7%
Debate	1	-	-	-	-	-	-
Problem Solving	6	-	-	2	12.5%	8	12.90%
Making Plans	3	-	-	1	6.25%	1	1.6 %
Simulations	6	-	-	-	-	-	-
Making Decisions	5	-	-	2	12.5%	9	14.5%
Narrative	6	1	1.6%	-	-	4	6.4%

Administrators Questionnaire: The third type of questionnaire (See Appendix C) was given to the three administrators at DBE so as to ascertain their ideas and suggestions in terms of content validity of the end-of-term OAT. There were only three questions besides the “comments” part in the questionnaire.

Question 1 asked whether they were generally pleased with the content of the OAT. It was a YES-NO type of question. Of the three administrators, two of them said they were not pleased and marked the “NO” option; however, one of them responded this question as “YES”.

Question 2 was for the one who responded to Question 1 as “YES”. It asked what the positive aspects of the current OAT were. In defending the OAT, the administrator said that students were familiar with the topics. In the comments part the administrator said “ Although my response is positive, I still believe that adaptations would be welcomed for better results”.

Question 3 was for those who responded to Question 2 as “NO”, which meant that they were not pleased with the content of the OAT. The two administrators who chose this option explained their reasons for that choice. First of all, they both stated that more variety in the speaking task types should be provided. As for the suggestions to improve the OAT, they said role plays, simulations, or making up stories following a sequence of pictures, or cartoons could be added to the content of the OAT. In addition, they both emphasized the importance and necessity of sub-topics in order to elicit more detailed responses. One of them clarified the rationale behind it by saying “Sub-topics related to the main topic chosen should be prepared beforehand. For example, if the main topic is traffic, the sub topics might be “ What

are the causes of the traffic accidents? Or What measures should be taken to prevent the traffic accidents?” In the comments part, both of them complained about the time allocated to each student in the OAT.

This chapter has given the detailed responses to the questionnaires and to the analysis of task types both in the course book and the OAT study sheet. In chapter 5 I give my conclusions.

CHAPTER 5: CONCLUSIONS

Summary of the Study

The aim of this research study was to investigate the content validity and backwash effect of the end-of-term Oral Assessment Test (OAT) administered at Hacettepe University, Department of Basic English (DBE). The content validity of the end-of-term OAT was examined in terms of consistency between the test content and the content of the course book. Apart from the content validity of the OAT, I examined the backwash effect of the OAT, which is the effect of the OAT on teaching and learning.

In this study, three groups of subjects were included. They were the 14 B-level subject teachers and 2 testers, 62 B-level students who took the OAT and 3 administrators. Three types of materials were used for this research study. The first one was the course book, New First Certificate Masterclass, which was used to identify the types of speaking tasks taught in B-level courses. The second kind of material was the OAT study sheet (Appendix D), which contains 65 speaking topics students were held responsible for in the OAT. The third kind of material used in this study was questionnaires. Three different types of questionnaires were developed and given to the three groups of subjects.

For this study, the data were collected through the examination of both the course book content and the content of the OAT, and questionnaires as well. To analyze data, first, the types and frequencies of the speaking tasks in the course book were compared with those which students were held responsible for in the OAT. The

result was displayed in a table. Second, the questionnaires were analyzed by means of frequencies and percentages. Open-ended questions in the questionnaires were analyzed by putting them into categories according to recurring themes.

Discussion of Findings and Conclusions

This section discusses the findings of the study and draws conclusions based on the research questions outlined in Chapter 1. Each sub-section relates to one of the research questions. Where relevant, references to other reported research in the literature are presented.

The first research question was: To what extent does the end-of-term OAT represent the speaking items specified in the course book content for the preparatory students at Hacettepe University, Department of Basic English?

The results of the comparison between the course book content and the content of the end-of-term OAT in terms of speaking tasks show that a majority of the task types in the course book content do not exist in the OAT (See Table 1). Thus, it can be concluded that although a great variety of speaking tasks were practiced in the classroom during the term, students were held responsible for just a few types of speaking tasks. It is also notable that the speaking task types students were held responsible for in the OAT are not interactive tasks. On the contrary, they are informative and narrative tasks. In other words, these types of speaking tasks lend themselves to answers that are easy to memorize. Interactive tasks, for example, role play, picture discussion and information gap, were all excluded from the OAT.

Taeduck and Finch (1998) point out that oral tests must be true assessments of spoken abilities, rather than an indication of how well a student can produce well

memorized responses. Another striking point with respect to the end-of-term OAT is that it is an achievement test, which, according to Brown (1996) should assess how much of the course material or sets of skills are taught in a course and learnt by the student. However, this oral test seems to be ineffective in terms of testing what the course materials offer. As a result, I can conclude the OAT does not seem to be a good representative of the course material, and therefore, the degree of content validity of the OAT is considered to be low. Hughes (1990) stresses the importance of the degree of content validity, the more likely it is to be an accurate measure of what is supposed to measure. He also points out that “a test is said to have content validity if its content constitutes a representative sample of the language skills, structures etc. with which it is meant to be concerned” (p.22).

In addition, the results of the questionnaires given to the subjects, that is B-level subject teachers and testers and B-level students, support my conclusions above in terms of the low degree of the content validity of the end-of-term OAT. The teachers' and testers' perceptions about the task types occurring in the OAT, in comparison with the ones in the course book content appear to be in the same line with the results indicated in Table 1. Most of the teachers and testers agreed that a great majority of the speaking tasks in the OAT consisted of Description and Expressing Ideas and Opinions. As far as B-level students' perceptions are concerned, their questionnaire results are similar.

The second research question was: Does the end-of-term OAT affect classroom teaching and learning (backwash effect), and if so, how?

In order to reveal backwash effect of the OAT, the responses given to the related

parts of the questionnaires were analyzed. Question 5 of the subject teachers' and testers' questionnaire asked whether students put a lot of effort and time into performing well in the classroom on the speaking tasks which were not assessed the OAT. The majority of the respondents (62.5%) marked the choice of "yes, but to some extent" and a quarter of respondents favored the choice of "not very much." The following question, Question 6 was about the reason why students did not put a lot of effort and time into performing the speaking tasks in the classroom well. The results show that students know those tasks are not assessed and are of no value in terms of pass and fail degree in the OAT.

The results of the students' questionnaire supported the results of the questionnaire given to the teachers and testers in terms of backwash effect. For example, the result of Question 3 indicates that more than half of the students say they used only outside materials and ignored the course book for the OAT preparation. Question 4 was for those who responded to Question 3 as "No". Of the rest, most of them said they used mostly outside materials in addition to the course book. The results show that the course book did not contribute to their oral exam preparation very much.

The results of both the teachers and testers' and students' questionnaires show that the OAT has a negative backwash effect to some extent; that is, it seems students ignored some parts of the course book. That they ignored some parts of the course book may mean that the OAT affected teaching and learning in the classroom somewhat negatively

Content validity and backwash effect are interrelated matters. When a test has

little or lacks content validity, a harmful (negative) backwash effect is unavoidable. Hughes (1996) points out that if a test in which major areas identified in the specification are under-represented or not represented at all, such a test is likely to have a harmful backwash effect. Areas which are not tested are likely to become areas ignored in teaching and learning. Morrow (1990) supports Hughes' opinion by saying "a test is valid when it has good washback effect" (p.6).

Limitations of the Study

The first limitation of the study was the randomly selected 65 students. The 65 students may not represent all B-level students' attitudes and opinions. The second limitation was about the administrators' questionnaire in which the first question inquired whether they were pleased with the content of the OAT. They had only "Yes" or "No" options. The subsequent questions were to be answered according to the option they chose on the first question. However, for the sake of remaining impartial, the question could have been designed in a neutral way in order to elicit their opinions about both the positive and negative aspects of the OAT. The third limitation was about the definitions of the task types. The type of speaking tasks may have been interpreted according to the respondents' insights and background knowledge.

Implications

General Implications

This study can be useful for teachers and testers who are interested in speaking tests as I examined the content validity of an OAT from various perspectives. In addition, I examined backwash effect of an OAT, which may help teachers consider

the implications of an oral test in terms of learning and teaching in classrooms. I hope through this study, teachers, testers and administrators will become more aware of the classroom implications of an oral test. Furthermore, from the practical point of view, this study may help testers to revise the content of oral tests if they have similar problems in terms of the necessity of including the variety of speaking task types.

Institutional Implications

The results of this research study may pave the way for the redesign of the content of the OAT, which is currently limited to few speaking task types. It seems that it is necessary to add a greater variety of task types, for example discussions, role plays and picture discussions. In this way, the degree of consistency may be increased and a negative backwash effect could be prevented. In addition, the Oral Assessment Test might be administered at least twice a term in addition to the one administered at the end of the term.

Further Research

This study focused on the content validity and backwash effect of the end-of-term Oral Assessment Test (OAT). Therefore, this study did not take into consideration other types of validity of the OAT. Reliability is another aspect which is as important as validity. Further research might be done to explore other types of validity and reliability of the OAT. Another research study could be done on the way of the OAT's administration and scoring.

REFERENCES

- Alderson, J. C., Clapham, C. & Wall, D. (1995). Language test construction and evaluation. Cambridge: Cambridge University Press.
- Alderson, J. C. & Wall, D. (1993). Does washback exist? Applied Linguistics, 14, 115-128.
- Bachman, L. F. (1991). Fundamental considerations in language testing. Hong Kong: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Hong Kong: Oxford University Press.
- Brown, J. D. (1996). Testing in language programs. Upper Saddle River: Prentice Hall Regents.
- Carey, L. M. (1988). Measuring and evaluating school learning. Newton, Massachusetts: Allyn and Bacon, Inc.
- Finch, E. A. & Tae-duck, H. (1998). Oral testing and self-assessment- the way forward? The Internet TESL Journal, 2, 1-7.
- Gude, K. & Duckworth, M. (1996). Proficiency Masterclass. Hong Kong: Oxford University Press.
- Haines, S. & Stewart, B. (1996). New First Certificate Masterclass. Oxford: Oxford University Press.
- Heaton, J. B. (1988). Writing English language tests. Hong Kong: Longman Group UK Ltd.
- Henning, G. (1987). A guide to language testing. Cambridge, MA: Newbury House Publishers.

Hoekje, B. & Linnell, K. (1994). Authenticity in language testing: Evaluating spoken language tests for international teaching assistants. TESOL Quarterly, 28, 103-126.

Hughes, A. (1990). Testing for language teachers. Cambridge: Cambridge University Press.

Jakeman, V. & McDowell, C. (1996). Cambridge practice tests for IELTS 1. Cambridge: Cambridge University Press.

Johnson, K. & Johnson, H. (Ed.). (1998). Encyclopedic Dictionary of Applied Linguistics. Oxford: Blackwell Publishers Ltd.

Kitao, S. K. & Kitao, K. (1998). Testing communicative competence. The Internet TESL Journal, II, 1-4.

Madsen, H. S. (1983). Techniques in testing. Oxford: Oxford University Press.

Richards, J. & Platt, J. & Weber, H. (1989). Longman Dictionary of Applied Linguistics. Hong Kong: Longman Group (FE) Ltd.

Underhill, N. (1987). Testing spoken language. Cambridge: Cambridge University Press.

University of Cambridge Local Examination Syndicate. (1995). First Certificate in English. UCLES.

Weir, C. J. (1990). Communicative language testing. Trowbridge, Wiltshire: Prentice Hall International (UK) Ltd.

APPENDIX A

RESEARCH QUESTIONNAIRE CONCERNING THE VALIDITY AND
BACKWASH EFFECT OF THE END-OF-TERM ORAL ASSESSMENT TEST

Dear Colleagues,

I am an MA-TEFL graduate student at Bilkent University. I am doing a research on the validity and backwash effect of the End-of-term Oral Assessment Tests (OAT) administered at the Preparatory School of English at Hacettepe University. I am interested in your opinions concerning these tests. Your responses will help me a great deal with my research. Your responses will be kept strictly confidential. You do not have to give your name and no one will know your specific answers to these questions. I will be grateful if you would take a few moments to complete the questions.

Thank you,

Hasan Ösken

SECTION 1 : GENERAL INFORMATION ON B-LEVEL SUBJECT TEACHERS AND TESTERS

Put a cross(X) in the appropriate box.

1- I am working as a subject teacher teaching B-level students ()

I am working as a tester ()

2- I have been teaching for

() 1-5 years () 6-10 years () 11-15 years () 16-20 years

3- I have been working at Hacettepe University for

() 1-5 years () 6-10 years () 11-15 years () 16-20 years

4- I have been assigned to administrate the end-of-term OAT

() more than 11 times () 4-10 times () 1-3 times () never

SECTION 2 : In this section you are given information about the course content in terms of speaking skills.

The course book, New First Certificate Masterclass contains several structured speaking tasks which provide opportunities for the general development of speaking skills. These tasks are integrated throughout the units in the course book and they are parts of the course book syllabus which are regarded as the syllabus of the overall language program at the DBE, since The DBE has not got its own syllabus, but follows the syllabus of the selected course book's.

The types of speaking tasks in the course book content are listed below. Put a tick next to the tasks which are assessed in the end-of term OAT.

Role Play	_____
Role Play Interview	_____
Information Gap	_____
Picture Description and Discussion	_____
Description of Place, Object, People or Social Events	_____
Expressing Ideas and Opinions	_____
Discussion on a Discursive Topic	_____
Debate	_____
Problem solving	_____
Making plans	_____
Simulations	_____
Making decisions	_____
Narrative	_____

SECTION 3 : Regarding both the speaking tasks listed above, which exist in the course book content, and the tasks which students are held responsible for the pass and fail degree in the end-of-term Oral Assessment Test (OAT) , please circle the choice that applies to you using the scale below.

1. The end-of-term OAT represents all types of the speaking tasks specified above.

Strongly agree	Agree	Disagree	Strongly disagree
-------------------	-------	----------	----------------------

2. How much do the tasks which are assessed in the end-of-term OAT represent the whole course content?

Represent fully	Represent to some extent	Little representation	No representation
--------------------	-----------------------------	--------------------------	----------------------

3. Are the tasks which are included and assessed in the end-of-term OAT adequately enough to measure students' overall speaking performance?

Yes, adequately enough	Not adequately enough but to some extent	No, not at all
---------------------------	---	----------------

4. Are the tasks which you did **not** tick in Section 2 valuable in terms of gaining and improving speaking skills?

Very valuable	Valuable	Somewhat valuable	Not very valuable
---------------	----------	-------------------	-------------------

5. Remembering the tasks you did **not** tick in section 2, do you think students put a lot of effort and time into performing well in the classroom on those speaking tasks that will not be assessed and are of no value in terms of pass and fail degree in the end-of-term OAT?

Yes, they put a lot of effort	Yes, but to some extent	Not very much	Not at all
----------------------------------	----------------------------	------------------	------------

6. If your response “Yes, but to some extent” or “Not very much” to

Question 5, the reason is that students know that those task are not assessed and are of no value in terms of pass and fail degree in the end-of-term OAT. As a result, the end-of-term OAT effects the learning and teaching of speaking skills somewhat negatively.

Strongly
agree

Agree

Disagree

Strongly
disagree

SECTION 4 : Please give your own response to the following questions.

1. As an English language instructor are you pleased with the end-of-term OAT?

Why / why not?

2. What do you think could be done to improve the end-of-term OAT ?

3. Do you think that SPEAKING as a skill should be emphasized more in the classroom? If your answer is “yes”, please explain the reason how it could be done and why it should be done. If your answer “no”, please make your explanation.

Thank you very much for your kind interest

HASAN ÖSKEN

APPENDIX B

QUESTIONNAIRE FOR B-LEVEL STUDENTS

This questionnaire is part of a research project for MA-TEFL Program at Bilkent University. Its main purpose is to investigate the consistency between the course content and the content of the end-of-term Oral Assessment Test.

You are one of the participants who has been selected randomly to complete this questionnaire. The aim of this study is not to evaluate your instructor, nor the general instruction at the Preparatory School of English at Hacettepe University, but to obtain as much data as I can in order to offer better oral tests for next generations. That is the reason why your opinion is of high value.

There is no risk involved and all responses will be kept confidential. Please answer the questions honestly. Your participation will be greatly appreciated.

Thank you for taking the time to answer this questionnaire.

Hasan Ösken

English Language Instructor at Hacettepe University

SECTION 1 : GENERAL INFORMATION ON B-LEVEL STUDENTS

Instructions: Please put a tick in the appropriate box.

1. Gender

Male ()

Female ()

2. Age group

17-20 ()

21-25 ()

over 25 ()

SECTION 2 : QUESTIONNAIRE ON THE REFLECTION OF THE END-OF-TERM ORAL ASSESSMENT TEST AND THE COURSE BOOK STUDIED WITHIN ONE-YEAR ACADEMIC TERM.

Instructions : Put a tick in the appropriate box.

1. Did you find the OAT easy?

Quite easy ()

Somewhat easy ()

Not very easy ()

2. Did you do any special study for the end-of-term Oral Assessment Test in your classroom as a part of class activities with your teacher?

Yes ()

No ()

If your answer “YES” to Question 2, put a tick in the appropriate box.

3. While doing the Oral Assessment Test preparation study in our classroom, we **only** used outside materials(e.g. Topics) which our teacher brought in the classroom, and we ignored our course book.

YES ()

NO ()

If your answer “YES”, do not answer the 4th question. If your answer is “NO”, please answer the following question.

4. For the Oral Assessment Test preparation, we used both the **course book** and **outside materials**.

Yes, we used both of them **equally** ()

Yes, but we used **mostly** the course book ()

Yes, but we used **mostly** the outside materials ()

5. Did the course book contribute much to your oral exam preparation?

Yes, a lot () Somewhat () Not very much ()

6. The types of speaking tasks in the content of the course book are listed below.

Put a tick next to the tasks on which you were assessed in the end-of-term Oral Assessment Test you took.

Role Play	()
Role Play Interview	()
Information Gap	()
Picture Description and Discussion	()
Description of Place, Object, People or Social Events	()
Expressing Ideas and Opinions	()
Discussion on a Discursive Topic	()
Debate	()
Problem Solving	()
Making Plans	()
Simulations	()
Making Decisions	()
Narrative	()

Thank you very much for your kind interest

HASAN ÖSKEN

APPENDIX C

RESEARCH QUESTIONNAIRE CONCERNING THE CONTENT VALIDITY
 AND BACKWASH EFFECT OF THE END-OF-TERM ORAL ASSESSMENT
 TEST AT THE DEPARTMENT OF BASIC ENGLISH AT HACETTEPE
 UNIVERSITY

As you know, I am engaged in a study about the content validity of the end-of-term Oral Assessment Test (OAT) at Hacettepe University, I would greatly appreciate your input. At this time I am only examining the content of the OAT, not the way it is administered and scored. Could you please respond with your ideas concerning the overall test content and the variety of speaking task types included.

Hasan ÖSKEN

1-Are you generally pleased with the content of the OAT?

.....Yes No

2-If yes, what are the positive aspects of the current OAT?

3- If no, what could be done to improve it?

Comments:

APPENDIX D

OAT Study Sheet Containing 65 Topics Used in the OAT. (Total 8 pages)

OAT STUDY SHEET

APPENDIX D

THE TOPICS USED IN THE END-OF-TERM OAT FOR B-LEVEL STUDENTS

May 18, 1999

1. If the time travel were possible, which period and which part of the would you return or go to? Why? What would you hope to see? Expressing Ideas and Opinions
2. If you could escape from the pressures / problems of your life, where would you escape to? What would you be escaping from? Expressing Ideas and Opinions
3. Newspaper or T.V.? Which is better for finding out what is happening in the world? Expressing Ideas and Opinions
4. What would be the advantages and disadvantages of living or working abroad? What would you miss about your own country? Expressing Ideas and Opinions
5. Has anything sad, funny, embarrassing or exciting ever happened to you? How did you feel? Narrative
6. Why are computer games so popular among children? Do you think parents are right to be worried? Expressing Ideas and Opinions
7. "The main purpose of education is to prepare people for jobs". Do you agree? Expressing Ideas and Opinions

8. What do you think are some of the dangers that face human beings in their daily life?

Expressing Ideas and Opinions

9. Which part of the world would you most like to live in? Give your reasons?

Expressing Ideas and Opinions

10. What do you like and dislike about the climate of your country? How does the climate of a country affect people?

Expressing Ideas and Opinions

11. How are old people treated in our country?

Expressing Ideas and Opinions

12. What are the advantages and disadvantages of living up to the age of 100? (Would you like to be 100? What are your reasons?)

Expressing Ideas and Opinions

13. What is your main reason for watching T.V? What kind of programmes do you prefer?

Expressing Ideas and Opinions

14. What are the best and worst features of T.V. programmes in your country?

Expressing Ideas and Opinions

15. Who decides which T.V. programmes you and your family watch?

Expressing Ideas and Opinions

16. Can you think of any serious 20th century problems which should be solved immediately?

Expressing Ideas and Opinions

17. What qualities do you think people need to do these jobs? - judge, reporter, fire fighter, hairdresser, nurse, teacher. Descriptive
18. Can you describe someone you like / admire or dislike very much? Descriptive
19. What is passive smoking? How are people affected by passive smoking? Expressing Ideas and Opinions
20. "Children whose parents smoke are more likely to smoke themselves". Do you agree with this idea? Expressing Ideas and Opinions
21. What is capital punishment? How many arguments can you think of for and against it? Expressing Ideas and Opinions
22. "Drinking too much can lead to crime." Do you agree with this opinion? Expressing Ideas and Opinions
23. Do you think the role of prison should be to punish or to reform criminals? Expressing Ideas and Opinions
24. "Schools should have programmes which educate young people about the dangers of taking drugs." Do you agree? Expressing Ideas and Opinions
25. If you had as much money and time as you wanted, what kind of holiday would you choose to have? Expressing Ideas and Opinions

26. What are some of the special restrictions which govern the lives of people who are school boys / doctors/ prisoners / sports personalities / blind people / old people / babies / fashion models.

Expressing Ideas and Opinions

27. What effect does lack of sleep have on you?

Expressing Ideas and Opinions

28. What special days are celebrated in our country? When and how are they celebrated? Descriptive

29. What do you do to help you remember things? Have you ever been embarrassed or annoyed with yourself because you've forgotten to do something?

Expressing Ideas and Opinions

30. Do you think most of the successful people were born with a silver spoon in their mouth? (How do you feel about the people who were born with a silver spoon in their mouth?) Expressing Ideas and Opinions

31. What is the definition of a successful person for you? Descriptive

32. Do you think exams are good criteria to evaluate one's success?

Expressing Ideas and Opinions

33. In many countries, there are interesting customs for births, marriages, the seasons or good luck. What interesting customs do you know?

Expressing Ideas and Opinions

34. Are there any skills or abilities that men and women are particularly good at?

Expressing Ideas and Opinions

35. Do you agree with the idea that all barriers between countries should be lifted?

Expressing Ideas and Opinions

36. Shopkeepers shouldn't sell cigarettes to children under 16. Do you agree or disagree?

Expressing Ideas and Opinions

37. Which of these technological inventions has had the greatest effect on people's lives – the telephone, the computer or the T.V.?

Expressing Ideas and Opinions

38. Why do children of famous people often follow their parents into the same kind of work? Does this just happen with famous people?

Expressing Ideas and Opinions

39. Many people have several different images – a private image and public images. How many images do you have? How would you describe them?

Expressing Ideas and Opinions

40. What are the man-made environmental disasters and their effects on the environment? What can and should be done to prevent it happening?

Expressing Ideas and Opinions

41. Do you think acting is a natural talent or an acquired skill? Why do children of famous actors or actresses often follow their parents into the same kind of work?

Expressing Ideas and Opinions

42. What would you do to stop someone who is trying to commit suicide?

Expressing Ideas and Opinions

43. What is one of the best books / films you have read/ seen? What is special about it? Why do you like/ dislike it?

Expressing Ideas and Opinions

44. What makes you angry in relationships with other people?

Expressing Ideas and Opinions

45. What are five things you look for the most in your future husband / wife?

Descriptive

46. For what reasons do you think people tell lies?

Expressing Ideas and Opinions

47. Imagine you are going on holiday. What precautions would you take?

Expressing Ideas and Opinions

48. Do you think people's hairstyles and clothes reflect their personality?

Expressing Ideas and Opinions

49. "We should protect the rights of animals" Do you agree or disagree? What is the situation / treatment towards animals like in our country?

Expressing Ideas and Opinions

50. In your opinion, what would be the advantages and disadvantages of living or working abroad?

Expressing Ideas and Opinions

51. In what ways are we looking after the world, and in what ways are we destroying it? Expressing Ideas and Opinions
52. Do you think arranged marriages are a good idea? Would you like your parents to find a partner for you? Expressing Ideas and Opinions
53. What changes have there been in the roles of men and women in our country in the last 100 years? Expressing Ideas and Opinions
54. If you were of mixed nationality, which nationalities would you like to be? Why? Expressing Ideas and Opinions
55. Do you think T.V. has an influence on the way you behave or think? Expressing Ideas and Opinions
56. What could be the problems you may have if you had to share a place with a friend? Expressing Ideas and Opinions
57. How important do you think it is for young people to have their own bedroom? (location, size, decoration) Expressing Ideas and Opinions
58. Why are big cities getting more and more dangerous to live? Expressing Ideas and Opinions
59. "Yesterday is gone, tomorrow may never come". What do you understand from this sentence? Expressing Ideas and Opinions

60. "All that glitters is not gold". Do you agree?

Expressing Ideas and Opinions

61. "Education mirrors society." Do you agree?

Expressing Ideas and Opinions

62. Men are generally luckier (happier) than women in the world! Do you agree? Why or why not?

Expressing Ideas and Opinions

63. What is the stereotype Turkish man or woman? Give some examples.

Descriptive

64. What kind of things do people like doing when they retire in your country? What would you like to do when you retire?

Expressing Ideas and Opinions

65. How much and in what ways are people influenced by advertisements?

Expressing Ideas and Opinions